

演化樹重組與最大相似值之研究

The study of phylogenetic tree reconstruction and maximum likelihood

指導老師: 謝孫源

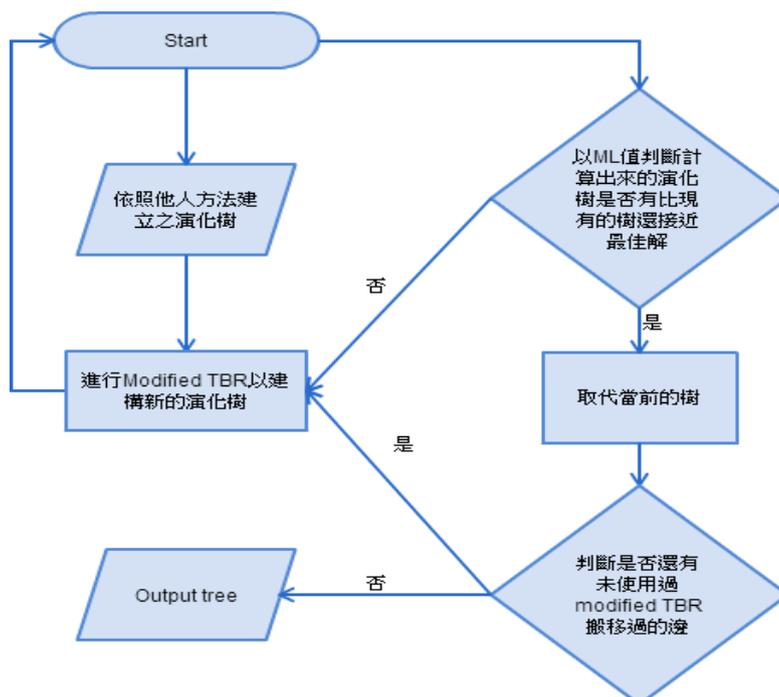
專題成員: 陳杰立

開發工具: C++

開發環境: Code Block

一、簡介

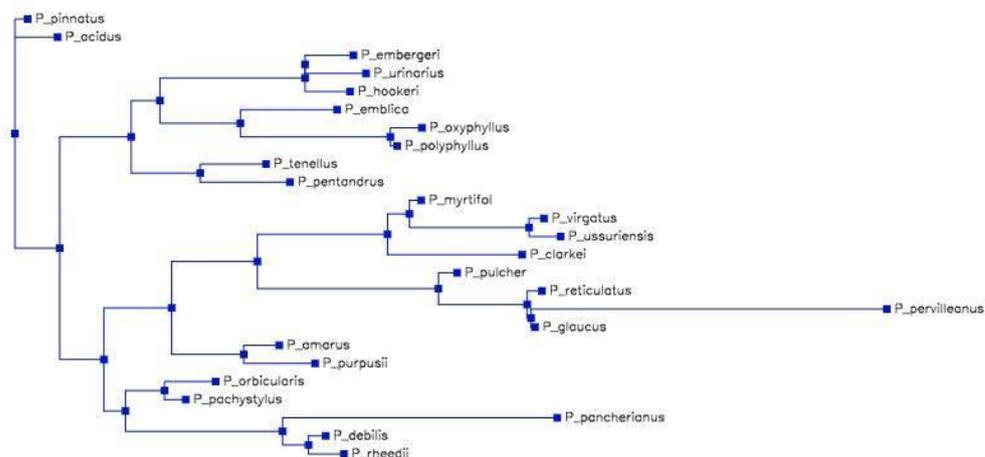
演化樹 (evolutionary tree)，是表明被認為具有共同祖先的各物種間演化關係的樹。是一種親緣分支分類方法 (cladogram)。在樹中，每個節點代表其各分支的最近共同祖先，而節點間的線段長度對應演化距離 (如估計的演化時間)。然而在建造演化樹的過程中可能會因為計算量太大而造成不必要的花費，所以我們要做的是如何把複雜度降低。tree bisection and reconnection(TBR)在重建的時候採用隨機選邊重建的方式，但是這種方式會多產出很多 Cost 較高的樹，增加時間複雜度，我們採用的 modify TBR 法是讀進他人方法所建構出來的演化樹並改進並且在每一次選取最大的邊來拆開重組，則可避免重建出多餘的高 Cost 樹來提高 TBR 的效能。此外為了增進所建構出來的樹的精準度，我們採用了 Minimum evolution 來刪除掉與估算差距過大的樹，然後再使用 Maximum likelihood 來確認是否新建的樹比原來的樹要來的好，以實現演算法的正確性與精準度。



二、測試結果

說明:

下圖為演化樹重建的結果，節點上顯示的為物種。而物種之間的關係越遠，邊長也會越長。



評比方法

Maximum Likelihood value

比較兩棵樹的相似程度，結果以 ML 值表示，值越接近 True Tree 算出來的 ML 值表示越接近真實的演化樹。

執行方式：

讀進他人方法(DNAML、PhyML、RAxML)所建構出來的演化樹(T_f)，並用我們的方法(DNAML+TRMLE、PhyML+TRMLE、RAxML+TRMLE)改進並產生新的演化樹(T_r)。比較 T_f 、 T_r 和 True Tree 計算出來的 ML 值(ML_f 、 ML_r 、 ML_t)。所得結果最貼近 ML_t 者為結果較好的演化樹。

測驗結果表:

物種數/方法	35	40	100
DNAML	-1003.66	-7045.72	-20485.7
PhyML	-1020.87	-6952.3	-20353.4
RAxML	-993.644	-7070.61	-20451.3
DNAML+TRMLE	-1018.25	-7105.5	-20932.8
PhyML+TRMLE	-1019.08	-7133.53	-20991.1
RAxML+TRMLE	-1020.65	-7130.71	-20956.2
Ture Tree	-1018.97	-7117.4	-20959.1

我們的方法是 TRMLE，其他人的方法是 DNAML、PROML、RAXML。
TRMLE 的 RF 值與其他方法比較數值都較小，所以我們的方法確實比其他人好。